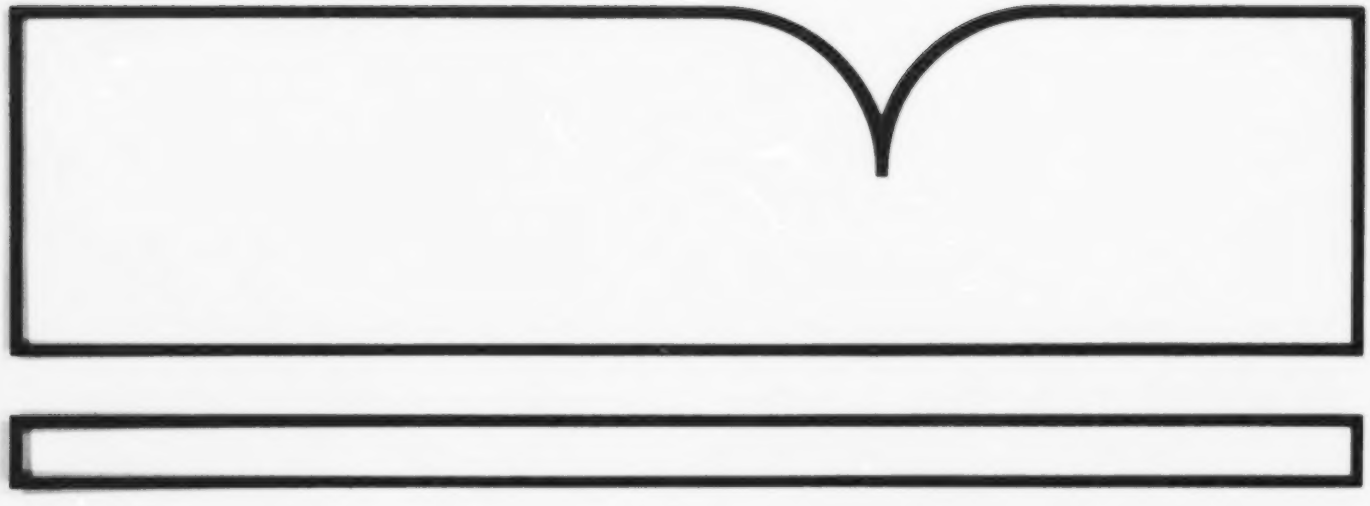
PB87-174876

Statistical Aspects of Autoregressive Models in the Assessment of Radon Mitigation

(U.S.) Environmental Protection Agency
Research Triangle Park, NC

Mar 87



U.S. Department of Commerce
National Technical Information Service
NTIS

PB87-174876

EPA/600/D-87/085
March 1987

STATISTICAL ASPECTS OF AUTOREGRESSIVE MODELS IN THE ASSESMENT OF RADON MITIGATION

by

James E. Dunn
University of Arkansas
Fayetteville, AR 72701

D. Bruce Henschel
Air and Energy Engineering Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

EPA Cooperative Agreement CR-812305-1-0

EPA Project Officer
Bruce A. Tichenor

AIR AND ENERGY ENGINEERING RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC 27711

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/D-87/085	2.
3. RECIPIENT'S ACCESSION NO. PB87 174876/AS	
4. TITLE AND SUBTITLE Statistical Aspects of Autoregressive Models in the Assessment of Radon Mitigation	5. REPORT DATE March 1987
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) James E. Dunn (University of Arkansas) and D. Bruce Henschel (EPA/AEERL)	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS The University of Arkansas 120 Ozark Hall Fayetteville, Arkansas 72701	10. PROGRAM ELEMENT NO.
	11. CONTRACT/GRANT NO. EPA Cooperative Agreement CR-812305-1-0
12. SPONSORING AGENCY NAME AND ADDRESS **EPA, Office of Research and Development** Air and Energy Engineering Research Laboratory **Research Triangle Park, NC 27711**	13. TYPE OF REPORT AND PERIOD COVERED Published paper; 6/85-12/86
	14. SPONSORING AGENCY CODE **EPA/600/13**

15. SUPPLEMENTARY NOTES AEERL project officer is Bruce A. Tichenor, Mail Drop 54, 919/541-2991. For presentation at 4th International Conference on Indoor Air, Berlin, Germany, August 1987.

16. ABSTRACT The paper discusses statistical aspects of autoregressive models in the assessment of radon mitigation. Radon values, as reflected by hourly scintillation counts, seem dominated by major, pseudo-periodic, random fluctuations. This methodological paper reports a moderate degree of success in modeling these data using relatively simple autoregressive-moving average models in order to assess the effectivess of radon mitigation techniques in existing housing. While accounting for the natural correlation of successive observations, familiar summary statistics such as steady state estimates, standard errors, confidence limits, and tests of hypothesis are produced. The Box-Jenkins approach is used throughout. In particular, intervention analysis provides an objective means of assessing the effectiveness of an active mitigation measure, such as fan off/on cycle. Occasionally, failure to declare a significant intervention has suggested a diagnosis of the data collection procedure.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Pollution	Pollution Control	13B
Radon	Stationary Sources	07B
Residential Buildings	Autoregressive Models	13M
Time Series Analysis	Intervention Analysis	12A

18. DISTRIBUTION STATEMENT **Release to Public**	19. SECURITY CLASS *(This Report)* **Unclassified**	21. NO. OF PAGES 12
	20. SECURITY CLASS *(This page)* **Unclassified**	22. PRICE

EPA Form 2220-1 (9-73)

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

STATISTICAL ASPECTS OF AUTOREGRESSIVE MODELS IN THE ASSESSMENT OF RADON MITIGATION

James E. Dunn
University of Arkansas, Fayetteville, Arkansas 72701, USA

D. Bruce Henschel
U.S. Environmental Protection Agency
Air and Energy Engineering Research Laboratory
Research Triangle Park, North Carolina 27711, USA

Abstract

Radon values, as reflected by hourly scintillation counts, seem dominated by major, pseudo-periodic, random fluctuations. This methodological paper reports a moderate degree of success in modeling these data using relatively simple autoregressive-moving average models in order to assess the effectiveness of radon mitigation techniques in existing housing. While accounting for the natural correlation of successive observations, familiar summary statistics such as steady state estimates, standard errors, confidence limits, and tests of hypothesis are produced. The Box-Jenkins approach is used throughout. In particular, intervention analysis provides an objective means of assessing the effectiveness of an active mitigation measure, such as a fan off/on cycle. Occasionally, failure to declare a significant intervention has suggested a diagnosis of the data collection procedure.

Introduction

Mitigation techniques currently are being compared by EPA/AEERL in a long term study of 60-70 existing houses in the Reading Prong area of eastern Pennsylvania, cf. Henschel and Scott (4). From a statistical viewpoint, analysis of the resulting radon levels, as reflected by hourly, standard pylon measurements, is challenging since the conclusions for the effectiveness of mitigation techniques must be reached while the inhabitants go about their usual daily activities; e.g., doors are opened and closed and clothes dryers and other heating units cycle on and off. Superimposed on this randomness is the venturi effect of wind on the structure, as well as a possible "stack" effect; i.e., the tendency for warm air to rise in the house. The net result typically is a time series with any central tendency about a natural steady state level greatly dominated by large, somewhat periodic, random fluctuations. Figure 1 is typical, reflecting 3 weeks of data for a house during which suction was being applied to a baseboard duct installation after the top wall voids had been mortared. Each (+) represents an hourly measured radon value in picocuries per liter (pCi/L). In metric units, 1 pCi = 0.037 Bq (becquerel).

Spectral analysis suggested a natural period of about 5 days for the data in Figure 1, but since no explanation could be given for such a deterministic result, this approach did not seem profitable to pursue. Neither did it seem worthwhile to attempt to fit other forms of deterministic models for the infusion of radon into homes [e.g., reference (6)] since random patterns seemed to be the dominant feature of the data.

Hence, to assess mitigation effectiveness in the presence of this uncertainty, we arrived at use of a class of empirical, stochastic models known as autoregressive-moving average (ARMA) models whose inferential basis was largely developed by Box and Jenkins (2). They possess attributes suitable to our requirements, namely: they are pseudo-periodic subject to stochastic variation; they compensate for the fact that successive hourly measurements are not likely to be statistically independent (ruling out use of much traditional statistical methodology); and they provide a basis for inference in familiar terms such as estimates of steady state, including standard errors and confidence intervals, and tests of significance of possible interventions such as a fan on/off condition, cf. Box and Tiao (3). Our purpose is to illustrate these features in a methodological paper which summarizes the behavior of Box-Jenkins techniques applied to 46 data sets obtained on five houses which were subjected to a variety of mitigation measures.

ARMA Models

A typical ARMA model to represent a time series $\{y_t\}$ of radon measurements is given by

$$y_t = (1-\phi_1-\phi_2)\mu+(1-\phi_1B-\phi_2B^2)\omega B^L x_t/(1-\delta B)+\phi_1 y_{t-1}+\phi_2 y_{t-2}+a_t-\theta_1 a_{t-1}, \quad (1)$$

where $\{x_t\}$ is possibly an associated intervention series, coded 0 or 1, and parameters μ (the steady state), ω, δ, ϕ_1, ϕ_2, and θ_1 are to be estimated from the data. The "white noise" series $\{a_t\}$ is assumed to consist of independent, $N(0,\gamma^2)$ random variables. Thus, the current observation, y_t, is modeled as the sum of the steady state level plus a weighted sum of the last two observations (representing system momentum) plus a new independent random impact, a_t, incurred from external sources during the current time period, plus some fraction of the previous random impact. The intervention series impacts on y_t to the extent that $\omega \neq 0$, and only if $x_t = 1$. The impulse response function, $\omega B^L/(1-\delta B)$, is expressed in terms of the backshift operator, B, where $Bz_t = z_{t-1}$, and $B^L x_t = x_{t-L}$. Thus, if $z_t = \omega B^L x_t/(1-\delta B)$ is the effect of the intervention, then by clearing the fraction and applying the backshift operators, one obtains $z_t = \delta z_{t-1} + \omega x_{t-L}$; i.e., the effect of the intervention is delayed L time units and then tapers to the ultimate gain of $\omega/(1-\delta)$. Aside from the intervention term, equation (1) defines an ARMA(P=1,2;Q=1) model, where P indicates the indices of the autoregressive terms and Q denotes those of the moving average terms. An analogous notation will be used to specify other alternative models.

Intuitively, the ARMA model reflects whatever pattern there is in the randomness of the observed time series. Given the correct form of the model, estimates of γ^2 and the coefficients in the model, and a normal random number generator, one could produce any number of time series simulations equivalent to that actually observed.

We have used the readily available SAS (5) procedure ARIMA to effect all model fits reported here, using the method of maximum likelihood. Model identification was accomplished by fitting a short menu of tentative models, then choosing among them by requiring first that the autocorrelations among the residuals not differ significantly from that expected of white noise, and second that the estimated coefficients be statistically significant. Both diagnostics are provided by the ARIMA procedure. Generally, adequate fits were obtained using relatively simple models. The frequencies of models required for the 46 data sets analyzed to date were: 14 AR(P=1), 7 AR(P=1,2), 5 AR(P=1,3), 1 AR(P=1,4), 3 ARMA(P=1;Q=1), 6 ARMA(P=1;Q=2), 6 ARMA(P=1,2;Q=1), 1 ARMA(P=1,4;Q=1), and 3 white noise.

An AR(P=1,3) model gave an adequate fit to the data shown in Figure 1. P-values related to lack of fit as reflected by autocorrelations among the residuals were 0.272, 0.559, and 0.661 for lags through 6, 12, and 18 hours, respectively. Estimates were $\mu = 72.1$ (s.e.=8.37), $\phi_1 = 1.12$ (s.e.=0.0298), and $\phi_2 = -0.140$ (s.e.=0.0301). T-statistics were 8.61, 37.40, and -4.65, respectively, and these are statistically significant when compared to standard normal deviates. The mean squared error (MSE), estimating γ^2, was 21.6 . One step ahead forecasts by the fitted model are plotted as the connected line in Figure 1. The estimated steady state (72.1) is plotted as a solid line, and 0.95 confidence limits ($72.1 \pm 1.96 \times 8.37$) for the true steady state are plotted as dashed (---) lines. By contrast, the arithmetic mean for the same data set was 75.6 . This small discrepancy between the sample mean and the efficient model estimate has held true for most data sets, the main difference being in the quality of the associated standard errors with the model estimate being preferred.

A word needs to be said about our stationarity assumption. Even though there is some visual impression of a decreasing trend over the 3 week period shown in Figure 1, we have not attemped to detrend the data. From the

physical nature of the situation, we feel that the house is immersed in a constant supply of radon, and that all fluctuations are caused by external random impacts, including those whose influence is seen only gradually.

Finally, a 0.99 upper tolerance limit is plotted as a dot-dash line (-·-) at 128 pCi/L in Figure 1. This is our best estimate of an upper bound which, if the physical conditions of the house remain the same, will be exceeded by radon values only 1% of the time. This value in general is given by $\mu + z_{0.99}\sigma$, where $z_{0.99} = 2.326$ is the upper 99th percentile of the standard normal distribution, and σ^2 is the variance of y_t. Clearly, an estimate for any other tolerance limit could be obtained by choice of the proper standard normal deviate.

Unfortunately, the ARIMA procedure does not list σ^2 since it is model dependent. However, by rewriting any ARMA model (aside from intervention) in matrix notation using an extension of a result given by Anderson (1); i.e.,

$$\underline{Y}_t = (1-\phi_1-\ldots-\phi_p)\mu\underline{e} + B\underline{Y}_{t-1} + T\underline{a}_t \ , \qquad (2)$$

where $\underline{Y}_t' = (y_1, y_2, \ldots, y_{t-p+1})$, $\underline{e}' = (1,0,\ldots,0)$, $\underline{a}_t' = (a_t, a_{t-1}, \ldots, a_{t-q})$, and B and T are appropriate matrices involving $\phi_1, \phi_2, \ldots, \phi_p$ and $\theta_1, \theta_2, \ldots, \theta_q$, respectively, then $\text{var}[\underline{Y}_t] = \Sigma = B\Sigma B' + \gamma^2 TT' + B\text{cov}[\underline{Y}_{t-1}, \underline{a}_t']T' + T\text{cov}[\underline{a}_t, \underline{Y}_{t-1}']B'$, from which σ^2 can be obtained as any of the diagonal elements of Σ. These are listed below for some of the more familiar models.

$$\text{AR(P=1):}\quad \sigma^2 = \gamma^2/(1-\phi_1^2) \ ,$$

$$\text{AR(P=1,2):}\quad \sigma^2 = \gamma^2(1-\phi_2)/[1-\phi_1^2(1+\phi_2)-\phi_2(1+\phi_2-\phi_2^2)] \ ,$$

$$\text{AR(P=1,3):}\quad \sigma^2 = \gamma^2[1-\phi_3(\phi_1+\phi_3)]/[1-\phi_1\phi_3(1+\phi_1^2+\phi_1\phi_3-\phi_3^2)-\phi_1^2-\phi_3^2(2-\phi_3^2)],$$

$$\text{AR(P=1,4):}\quad \sigma^2 = \frac{\gamma^2[(1-\phi_4)(1-\phi_4^2)-\phi_1^2\phi_4]}{(1-\phi_1^2-\phi_4^2)[(1-\phi_4)(1-\phi_4^2)-\phi_1^2\phi_4]-2\phi_1^2\phi_4[\phi_1^2+\phi_4(1-\phi_4)]} \ ,$$

$$\text{ARMA(P=1;Q=1):}\quad \sigma^2 = \gamma^2[1+\theta_1(\theta_1-2\phi_1)]/(1-\phi_1^2) \ ,$$

$$\text{ARMA(P=1;Q=2):}\quad \sigma^2 = \gamma^2[1+\theta_2(\theta_2-2\phi_1^2)]/(1-\phi_1^2) \ ,$$

$$\text{ARMA(P=1,2;Q=1):}\quad \sigma^2 = \frac{\gamma^2\{(1-\phi_2)[1+\theta_1(\theta_1-2\phi_1)]-2\phi_1\phi_2\theta_1\}}{1-\phi_1^2(1+\phi_2)-\phi_2(1+\phi_2-\phi_2^2)} \ .$$

MSE from the model fit may be substituted for γ^2 and maximum likelihood estimates substituted for the remaining coefficients in order to evaluate σ^2. Using the formula corresponding to AR(P=1,3), $\sigma^2 = 631$ for the data in Figure 1, compared to the familiar sample variance $s^2 = 597$. In general, s^2 tends to underestimate the true variance.

Intervention analysis

Table 1 shows results of a variety of interventions in three different houses. In each case, the lag shown represents the most probable lag as reflected by a cross correlation analysis between the radon series and the "off-on" intervention series. Lack-of-fit p-values correspond to the model listed, and reflect the failure of autocorrelations up to lag 6 of the residual series to resemble that for white noise. With the possible exception of House 3, the models gave excellent fits.

The effect of an automatic air-to-air heat exchanger (AAHX) in the pres-

ence of sub-slab suction ventilation was being tested in House 2A. In the second run, the result was as expected, the estimated steady state radon concentrations changing from 565 to 201 pCi/L for a statistically significant gain of -364. For the first run, even though the estimated gain was -436 pCi/L, the p-value was 0.162; i.e., odds of 16 in 100 of falsely concluding an effect of AAHX. Figure 2 shows the problem; there simply was too much variation during the AAHX off period. In retrospect, we feel that this may have been introduced by entries and exits of technicians during the AAHX off period.

Table 1. Significance of interventions on radon levels in existing houses.

HOUSE/MODEL/DESCRIPTION	P FOR FIT	BASE LEVEL	INTERVENTION	LAG	GAIN	P FOR GAIN
		(pCi/L)	(pCi/L)		(pCi/L)	
2A.1: AR(P=1,2). One 250 cfm (0.118 m^3/s) suction fan ventilating sub-slab and one wall; slab sealed.	0.715	AAHX off 842	AAHX on 406	0	-436	0.162
2A.2: AR(P=1,2;Q=1). Above suction fan on; fireplace wall sealed with one suction point.	0.520	AAHX off 565	AAHX on 201	0	-364	0.001
3: AR(P=1,2). One 250 cfm suction fan ventilating sub-slab.	0.087	Fan off 1128	Fan on 370	3	-758	0.159
5: AR(P=1). Wall ventilation off; top voids sealed. Coal stove off/on in basement.	0.478	Stove off 46.2	Stove on 104	0	57.4	0.001

House 3 illustrated a problem of a different kind. Even though sub-slab suction resulted in an estimated 758 pCi/L reduction in steady state radon levels, again the odds are about 16 in 100 of falsely concluding an effect. Figure 3 reveals the data problem. There simply wasn't enough follow-up under the fan-on condition to accurately determine its steady state. The long drop to 370 pCi/L could readily be explained as simply a continuation of the downswing of the last cycle under the fan-off condition.

Figure 4 for House 5 shows a significant increase in the radon level of the basement as a result of a coal stove being lit. Though not shown in Table 1, the same highly significant increase was detected upstairs (with no apparent delay as reflected by cross-correlation analysis), the steady state estimate moving from 50.9 to 95.4 pCi/L for a gain of 44.5 (p=0.001).

Even though a p-value seems esoteric compared to ducts, drains, fans, and mortar, we believe that it has provided us with a means of being objective in our assessment of the effects of mitigation measures. One too easily assigns meaning to natural fluctuations of a radon series. When the observed change has seemed significant but the formal test has not, we usually discover that the problem is due to inadequacies in the data, as illustrated by two of the previous examples.

References

1. Anderson, T.W. The Statistical Analysis of Time Series. New York: John Wiley, 1958, pp. 177-178.

2. Box, G.E.P., and Jenkins, G.M. Time Series Analysis: Forecasting and Con-

trol. San Francisco:Holden-Day, 1970.

3. Box, G.E.P., and Tiao, G.C. Intervention analysis with applications to economic and environmental problems. American Statistical Asso. J. 70 (1975), 70-79.

4. Henschel, D.B., and Scott, A.G. The EPA program to demonstrate mitigation measures for indoor radon: Initial results. Proc. of an APCA International Specialty Conference (Philadelphia, February 1986), pp. 110-121.

5. SAS/ETS User's Guide. Cary, NC:SAS Institute Inc., 1982, pp. 77-113.

6. U.S. Department of Energy. Indoor Air Quality Environmental Information Handbook. Radon Report DOE/PE/72013-2, 1986.

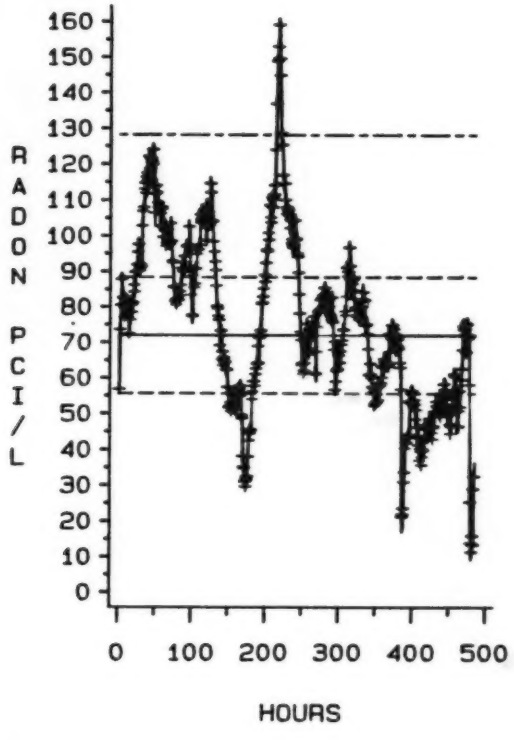


Fig. 1. Typical radon variation in the basement of a house, showing the steady state estimate (——), 0.95 confidence limits for the true steady state (– – –), and a 0.99 upper tolerance limit (–·–).

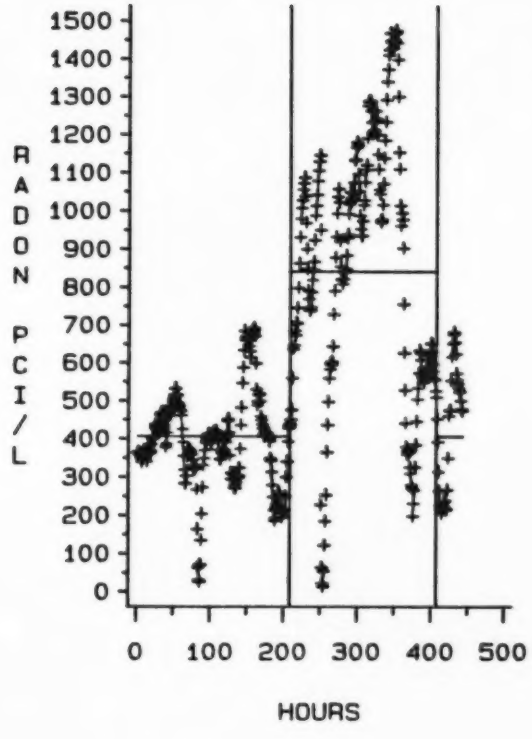


Fig. 2. Radon variation as an air-to-air heat exchanger cycled on/off/on. (——) = steady state; (|) = time of intervention.

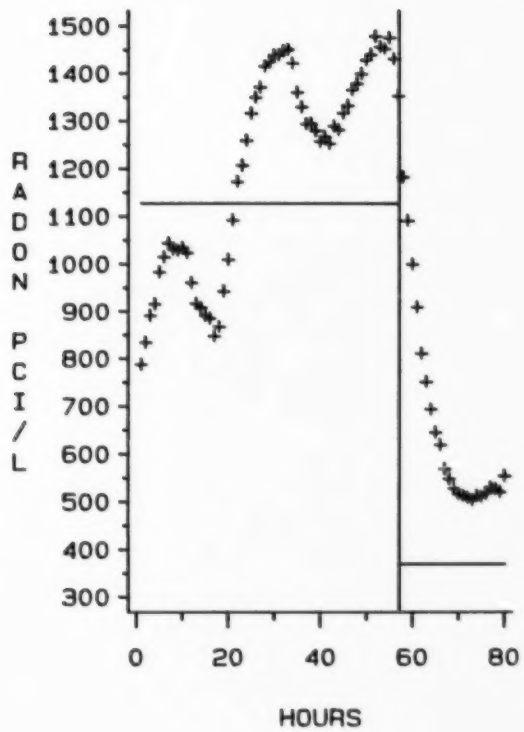


Fig. 3. Radon variation as sub-slab suction cycled off/on. (——) = steady state; (|) = time of intervention.

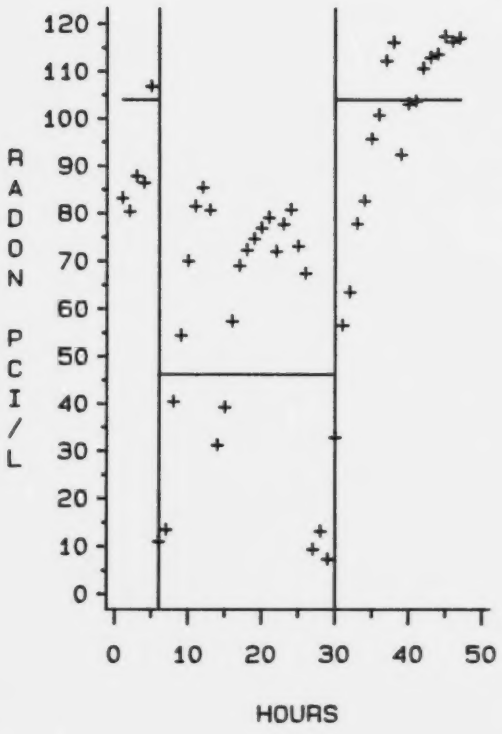


Fig. 4. Radon variation as a basement coal stove cycled on/off/on. (——) = steady state; (|) = time of intervention.